

# Blind Calibration of Mobile Sensors Using Informed Nonnegative Matrix Factorization

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel

LISIC, ULCO, Université Lille Nord de France, Calais, France  
firstname.lastname@lisic.univ-littoral.fr

**Abstract.** In this paper, we assume several heterogeneous, geolocalized, and time-stamped sensors to observe an area over time. We also assume that most of them are uncalibrated and we propose a novel formulation of the blind calibration problem as a Nonnegative Matrix Factorization (NMF) with missing entries. Our proposed approach is generalizing our previous informed and weighted NMF method, which is shown to be accurate for the considered application and to outperform blind calibration based on matrix completion and nonnegative least squares.

**Keywords:** Blind calibration, mobile sensor network, informed nonnegative matrix factorization, missing values

## 1 Introduction

Monitoring a natural or an industrial area is usually obtained from automated measurements provided by a set of sensors or from campaigns conducted by scientists. In the first case, the collected data are very accurate but the high sensor cost limits their number, hence a poor spatial sampling rate over the area. In the second case, the geographical coverage is large but—as the cost of such campaigns is high—sustaining them is difficult, hence a very low time sampling rate. Wireless Sensor Networks (WSN) were shown to solve the drawbacks of both above strategies: sensors in WSN are usually cheap and mobile, thus allowing their massive deployment for both an accurate time and spatial sampling. Moreover, even if they are individually less accurate than high cost sensors, they globally provide a similar accuracy while adding a finer spatial resolution [6]. However, their calibration is an issue—as the sensors might not be accessible—and blind sensor calibration techniques were proposed in, e.g., [1, 4, 9, 11, 12, 14, 16, 18] for that purpose. These methods may be divided into two categories, depending if the sensor network is mobile [9, 12, 14, 18] or not [1, 4, 11, 16]. Blind Mobile Sensor Calibration (BMSC) techniques usually assume that sensors in the same vicinity should provide the same data while Blind Fixed Sensor Calibration (BFSC) methods need additional assumptions about the acquired signals in order to perform the calibration, namely the measurement matrix in a compressed sensing framework [4, 16] or the low-rank subspace in which the observed data lie [1, 11].

In this paper, we investigate the BMSC problem as an informed matrix factorization, since our formulation provides a specific structure of the matrix factors. Assuming the calibration parameters, the acquired signals, and the sensed

physical phenomenon to be nonnegative, we revisit blind calibration as an informed Nonnegative Matrix Factorization (NMF) problem that we solve with an extended version of our previous work [10].

The remainder of the paper is structured as follows. We introduce the considered problem in Section 2, for which we propose a solution in Section 3. Section 4 investigates the experimental performance of our proposed method while we conclude and discuss about future work in Section 5.

## 2 Problem Statement

In this paper, we assume that a geographical area is observed by  $m$  heterogeneous, geolocalized, time-stamped, and mobile sensors along time. Such a situation arises in crowdsensing for example [8], where volunteers share some sensed information provided by their mobile device, e.g., their smartphone. The data obtained with crowdsensing are usually irregularly sampled in both the time and the space, hence the need of appropriate methods to process them. In this paper, we focus on blindly calibrating the sensors, for which we first introduce the definitions and assumptions used in this paper.

**Definition 1 ([15]).** *A rendezvous is a temporal and spatial vicinity between two sensors.*

Sensors in rendezvous should acquire the same phenomenon, thus providing the same data. Such rendezvous are classically used in BMSC. Most approaches—e.g., [9,12]—consider that uncalibrated data are randomly distributed around the calibrated ones, so that averaging the measurements in rendezvous performs the calibration. However, the averaging-based calibration is not always applicable—see [7, Fig. 2] for example. As an alternative, some authors consider rendezvous between both calibrated and uncalibrated sensors [14] in order to locally perform the calibration<sup>1</sup>. The newly calibrated sensors are then used to calibrate the still uncalibrated ones in other rendezvous and so on. Such a multi-hop calibration technique needs a dense network to be deployed, so that one can ensure each sensor to be in rendezvous with a (newly) calibrated sensor [14]. However, multi-hop calibration might suffer from propagating calibration estimation errors.

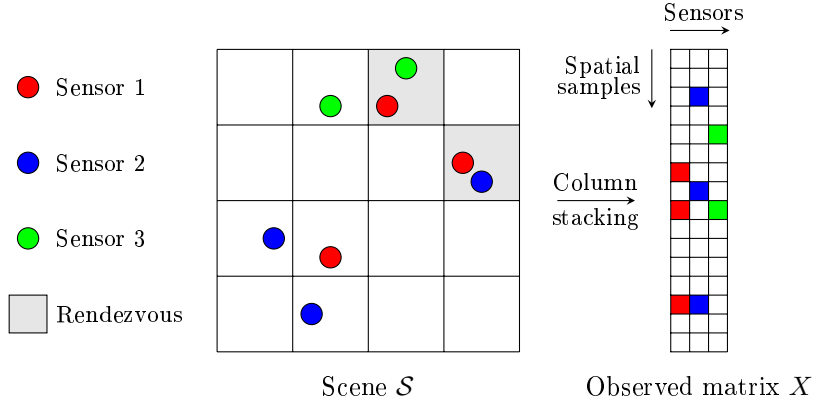
In this paper, we start from the same idea of rendezvous and the same hypothesis of dense sensor network but we propose a matrix formulation allowing to calibrate the *whole* sensor network without multiple hops. For that purpose, we first introduce the following definition.

**Definition 2.** *A scene  $\mathcal{S}$  is a discretized area observed during a time interval  $[t, t + \Delta t)$ . A spatial pixel has a size lower than  $\Delta d$ , where  $\Delta t$  and  $\Delta d$  define the vicinity of the rendezvous.*

A scene can thus be seen as a grid of locations where sensors go to and where they sense a physical phenomenon. When two sensors share a common position

<sup>1</sup> Using calibrated and uncalibrated sensors has also been considered in BFSC [1, 11].

in a scene, they are in rendezvous (Fig. 1). Setting  $\Delta t$  and  $\Delta d$  in order to define a scene highly depends on the nature of the sensed phenomenon [15]. In this paper, we assume to observe one scene<sup>2</sup>  $\mathcal{S}$  that we rearrange as a vector  $\underline{y} \triangleq [y(1), \dots, y(n)]^T$ , where  $n$  is the number of space samples in  $\mathcal{S}$ .



**Fig. 1.** From a scene  $\mathcal{S}$  (with  $n = 16$  spatial samples,  $m = 3$  sensors and 2 rendezvous) to the data matrix  $X$  (white pixels mean no observed value).

We now assume that  $m$  heterogeneous sensors are observing  $\mathcal{S}$  (see Fig. 1). Let  $x(i, j)$  be the sample from Sensor  $j$  corresponding to the  $i$ -th sample in  $\underline{y}$ . Physically,  $x(i, j)$  is a sensor-output voltage which is here assumed to be linked to the sensed phenomenon  $y(i)$  according to an affine relationship, i.e.,

$$x(i, j) \simeq y(i) \cdot \alpha_j + \beta_j, \quad (1)$$

where  $\alpha_j$  and  $\beta_j$  are the unknown gain and offset associated with Sensor  $j$ , respectively. These coefficients are assumed to be constant over the scene<sup>3</sup>.

We now define  $G$  and  $F$ , the  $n \times 2$  and  $2 \times m$  matrices which respectively read

$$G \triangleq \begin{bmatrix} y(1) & 1 \\ \vdots & \vdots \\ y(n) & 1 \end{bmatrix} \text{ and } F \triangleq \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ \beta_1 & \beta_2 & \cdots & \beta_m \end{bmatrix}. \quad (2)$$

If we assume each sensor to observe the whole scene  $\mathcal{S}$ , the matrix form of Eq. (1) then reads

$$X \simeq G \cdot F, \quad (3)$$

where  $X \triangleq [x(i, j)]$  is a low-rank matrix, assumed to be well-conditioned. Solving the blind calibration problem then consists of estimating  $F$  from the observed matrix  $X$ . In particular—considering the  $n \times m$  weight matrix  $W$  defined as

$$W(i, j) \triangleq \begin{cases} 0 & \text{if } x(i, j) \text{ is not available,} \\ \rho_j & \text{otherwise,} \end{cases} \quad (4)$$

<sup>2</sup> The case of multiple scenes—out of the scope of this paper—is discussed in Section 5.

<sup>3</sup> Some authors, e.g., [9], consider the sensor responses to drift over time.

where  $\rho_j$  is a weight coefficient associated with Sensor  $j$ —BMSC for Scene  $\mathcal{S}$  can be written as a weighted matrix factorization problem, i.e.,

$$\min_{G \geq 0, F \geq 0} \|W \circ (X - G \cdot F)\|_f^2, \quad (5)$$

where  $\|\cdot\|_f$  is the Frobenius norm and where  $\circ$  denotes the Hadamard product.

In this paper, we assume  $X$ ,  $G$ , and  $F$  to be nonnegative, which is satisfied in practice for several environmental sensors such as dust sensors [17]. We now introduce our proposed weighted NMF method for estimating  $F$ .

### 3 Proposed Blind Calibration Method

In this section, we introduce our proposed approach for BMSC. The problem (5) may yield scale ambiguities in the columns of  $F$ , as for any source separation problem. In order to solve them, we consider that one sensor—say Sensor  $m$ —is calibrated and that its calibration parameters are respectively equal to<sup>4</sup>

$$\alpha_m = 1 \quad \text{and} \quad \beta_m = 0. \quad (6)$$

i.e., if the value  $x_{im}$  is available, we get  $x_{im} \simeq y_i$ . At this stage, it should be noticed that the factorization problem in Eq. (5) is informed. Indeed, Eqs. (2) and (6) show the last column in both  $G$  and  $F$  to be known. Taking into account such information should improve the factorization and fix the scale ambiguity inherent to blind factorization.

Recently, informed NMF methods were proposed [5, 10] and considered the available information as a penalization term in the NMF optimization [5] or as a specific parameterization which sets the known parameters [10]. As a penalization term does not freeze the known entries of  $G$ —which do not depend on the observed data—the latter strategy seems better suited for the considered problem<sup>5</sup>. However, the parameterization in [10] only considers information on  $F$ . In this paper, we thus generalize [10] in order to apply it to the considered BMSC problem.

Using the same formalism as in [10], we define  $\Omega^F$  and  $\Omega^G$ , the binary matrices which inform the presence/absence of constraints in  $F$  and  $G$ , respectively. We then define  $\Phi^F$  and  $\Phi^G$ , the sparse matrices of set entries in  $F$  and  $G$ , respectively. It should be noticed that

$$\Phi^F = \Omega^F \circ F \quad \text{and} \quad \Phi^G = \Omega^G \circ G. \quad (7)$$

Defining  $\overline{\Omega}^F \triangleq 1 - \Omega^F$  (respectively  $\overline{\Omega}^G \triangleq 1 - \Omega^G$ ) and  $\Delta^F$  (respectively  $\Delta^G$ )—the matrix of free parameters in  $F$  (respectively in  $G$ )—we extend [10]

<sup>4</sup> Actually, if no calibrated sensor is available, it is still possible to perform a *relative* calibration, thus providing some consistency in the sensor responses [1, 4, 11].

<sup>5</sup> An alternative might consist of successively (i) updating  $F$  or  $G$  with usual update rules and (ii) replacing the known entries by their actual values at each iteration. However, this strategy yielded a low performance in some preliminary tests.

and derive

$$F = \Omega^F \circ \Phi^F + \overline{\Omega}^F \circ \Delta F \quad \text{and} \quad G = \Omega^G \circ \Phi^G + \overline{\Omega}^G \circ \Delta G. \quad (8)$$

Solving Eq. (5) is then performed using an alternating technique, where we successively aim to estimate  $F$  and  $G$  using the parameterization in Eq. (8), i.e.,

$$\min_{\Delta F \geq 0} \|W \circ (X - G \cdot \Phi^F - G \cdot \Delta F)\|_f^2 \quad \text{and} \quad \min_{\Delta G \geq 0} \|W \circ (X - \Phi^G \cdot F - \Delta G \cdot F)\|_f^2. \quad (9)$$

Using the same strategy as in [10]—the proof is omitted for space considerations but it is based on a Majoration-Minimization optimization—we derive the update rules which read, respectively:

$$F \leftarrow \Phi^F + \Delta F \circ \overline{\Omega}^F \circ \left[ \frac{G^T (W \circ (X - G \cdot \Phi^F)^+)}{G^T (W \circ (G \cdot \Delta F))} \right], \quad (10)$$

$$\text{and } G \leftarrow \Phi^G + \Delta G \circ \overline{\Omega}^G \circ \left[ \frac{(W \circ (X - \Phi^G \cdot F)^+) F^T}{(W \circ (\Delta G \cdot F)) F^T} \right]. \quad (11)$$

The superscript  $+$  here denotes the function defined as  $(z)^+ \triangleq \max\{\epsilon, z\}$ , where  $\epsilon$  is a small user-defined threshold. It should be noticed that, contrary to [10], we here consider a matrix factorization problem with missing entries. Let us remind that calibration can only be done if the sensor network is dense enough, as assumed in Section 2.

Lastly, in order to apply informed NMF to the considered problem, we must initialize  $G$  and  $F$ , which is known to be tricky. Classical strategies consist of a random initialization while some authors propose an initialization provided by experts [10], the output of another factorization method [3], or a physical model [13]. In this paper, we take advantage of the fact that  $X$  is low rank and that some entries of  $X$  are equal some of  $G$  according to Eq. (6). Our proposed strategy to initialize NMF then consists of applying a matrix completion technique [2] to  $X$ . The completed matrix is denoted  $\tilde{X}$  hereafter. By construction, the last column of  $\tilde{X}$  is an estimation of the first column of  $G$ . Estimating  $F$  can then be obtained from  $\tilde{X}$  and  $G$  using nonnegative least-squares. In this paper, we study the enhancement provided by our proposed informed NMF with respect to the matrix-completion-based calibration, that we use for initializing NMF.

## 4 Experimental Validation

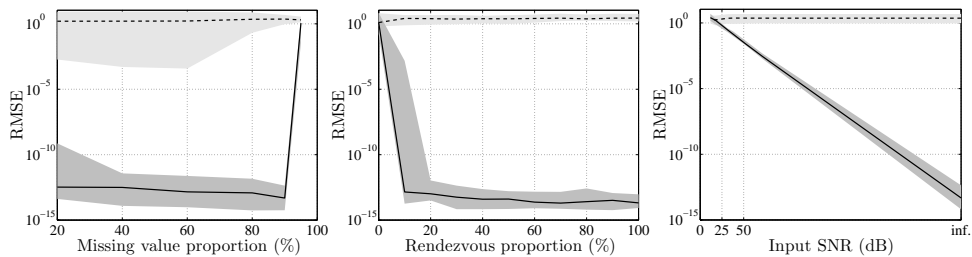
In this section, we aim to investigate the enhancement provided by our proposed informed NMF method for BMSC. For that purpose, we simulate a crowdsensing-like particulate matter sensing during a time interval  $[t, t + \Delta t)$ , which satisfies the assumptions in Section 2. The scene is a  $10 \times 10$  discretized area (the length of  $\underline{y}$  is thus equal to  $n = 100$ ) which is observed by  $m = 26$  sensors, i.e.,  $m - 1$  uncalibrated and mobile dust sensors [17] connected to mobile devices and one calibrated, high quality, and mobile sensor<sup>6</sup>.

<sup>6</sup> Actually, we get  $k$  fixed, calibrated, and accurate sensors whose obtained values are modeled as those of the  $m$ -th sensor in the BMSC problem.

The observed concentrations in  $\underline{y}$  range between 0 and 0.5 mg/m<sup>3</sup>, for which the sensor response is assumed to be affine [17]. For each uncalibrated sensor, each observed data point represents a nonnegative voltage linked to the corresponding ground truth point in  $\underline{y}$  according to Eq. (1). In particular, following the datasheet in [17], the gain and offset coefficients  $\alpha_j$  and  $\beta_j$  are randomly set according to a Gaussian distribution centered around 5 V/(mg/m<sup>3</sup>) and 0.9 V, respectively, and then projected onto their respective interval of admissible values—provided by the manufacturer [17]—i.e.,  $3.5 < \alpha_j < 6.5$  and  $0 < \beta_j < 1.5$ ,  $\forall j = 1, \dots, m-1$ . We then get a  $26 \times 100$  theoretical observation matrix for which we randomly keep  $k+l$  samples in  $X$  only, where  $k$  (respectively,  $l$ ) is the number of calibrated (respectively, uncalibrated) sensor samples—with  $k \ll l$ —hence providing the irregular spatial sampling over the scene. Lastly, Gaussian noise realizations may be added to the observed uncalibrated sensor data and the weight coefficients  $\rho_j$  defined in Eq. (4) are set to

$$\rho_j = 1, \forall j = 1, \dots, m-1, \quad \text{and} \quad \rho_m = l. \quad (12)$$

In this section, we aim to explore the influence of the number of rendezvous between calibrated and uncalibrated sensors, the number of missing entries in  $X$  and the influence of the input SNR to the BMSC performance. For each test condition—i.e., one number of rendezvous, one proportion of missing entries, or one input SNR—25 simulations are performed. In each run, we randomly set the positions of the samples in  $X$  in the three experiments and we generate different noise realizations in the last one. The number  $k$  of calibrated sensor values in the  $m$ -th column of  $X$  is set to  $k = 4$  in all the tests. Except when we make these values vary, the proportion of uncalibrated sensors to have rendezvous with calibrated ones, and the proportion of missing entries in  $X$  are set to 30% and 90%, respectively.



**Fig. 2.** Performance of the BMSC method vs (*left*) the proportion of missing values, (*center*) the proportion of rendezvous between calibrated and uncalibrated sensors, and (*right*) the input SNR.

Figure 2 shows the Root-Mean Square Error (RMSE) achieved by our proposed method in the above test configurations. The RMSEs are computed over

the first line<sup>7</sup> of  $F$ , for the uncalibrated sensors only. Dark gray (respectively, light gray) areas show the RMSE envelope while the solid lines (respectively, dashed lines) represent the median RMSE obtained after  $10^6$  NMF iterations (respectively, at initialization). On the left plot, we first investigate the effect of the proportion of missing entries—ranging from 20 to 95%—on BMSC in a noiseless configuration. Our proposed method is robust to the number of missing entries (with consistent RMSEs below  $10^{-10}$  until 90% of missing values) and outperforms the matrix-completion-based initialization (with a median RMSE around 3). When 5% of the data in  $X$  are available, the dense network assumption is not satisfied anymore and the performance drastically decreases.

The central plot shows the influence of the number of uncalibrated sensors to have rendezvous with calibrated ones. The achieved performance is quite similar to the previous one, except that the upper side of the dark gray envelope is much higher than the median RMSE when the rendezvous proportion is equal to 10%. Please note even when the calibration error is high, rows of  $F$  are correctly estimated, up to a scale coefficient which cannot be handled anymore.

The right plot shows the influence of the input SNR on the BMSC performance. In addition to the noiseless case, we make vary the SNR from 11 to 70 dB. The calibration accuracy decreases with the input SNR and is quite similar to the one obtained with matrix completion for the lowest tested input SNRs.

## 5 Conclusion and Discussion

In this paper, we revisited blind mobile sensor calibration as a matrix factorization problem. Assuming any of the matrices in the factorization to be non-negative, we generalized our previous informed NMF [10] for the considered application. The approach was shown to be robust to the number of missing entries and to the number of rendezvous between calibrated and uncalibrated sensors. However, some assumptions—e.g., the dense sensor network over the considered zone, or the fact that  $X$  is well-conditioned—might seem restrictive. It should be noticed that the approach proposed in this paper can be extended to the case of multiple scenes, by stacking all the observed—and sufficiently different—matrices in one unique well-conditioned matrix, so that we multiply the number of both the known sensors and the rendezvous. In that case, it is more likely that the calibration assumptions will be satisfied.

In future work, we aim to explore several directions. As mentioned above, joint-factorization will be investigated. Moreover, the NMF method proposed in this paper is an extension of the Lee and Seung multiplicative update algorithm, which is known to be slow to converge when the size of the data matrix is large. Extending recent and fast NMF methods to our informed framework will be considered. We will also explore the calibration enhancement provided by some spatial information about the scene  $\mathcal{S}$  and the effects of the scene spatial discretization on the calibration performance.

---

<sup>7</sup> RMSEs computed over the second line of  $F$ —not shown for space consideration—are similar to those plotted in Fig. 2.

**Acknowledgments.** This work was funded by the “OSCAR” project within the Région Nord – Pas de Calais “Chercheurs Citoyens” Program.

## References

1. Balzano, L., Nowak, R.: Blind calibration of sensor networks. In: Proc. of IPSN. pp. 79–88 (2007)
2. Becker, S., Candès, E., Grant, M.: Templates for convex cone problems with applications to sparse signal recovery. *Math. Progr. Comp.* 3(3), 165–218 (2011)
3. Benachir, D., Deville, Y., Hosseini, S., Karoui, M.S., Hameurlain, A.: Hyperspectral image unmixing by non-negative matrix factorization initialized with modified independent component analysis. In: Proc. of WHISPERS (2013)
4. Bilen, C., Puy, G., Gribonval, R., Daudet, L.: Convex optimization approaches for blind sensor calibration using sparsity. *IEEE Trans. on Signal Processing* 62(18), 4847–4856 (Sept 2014)
5. Choo, J., Lee, C., Reddy, C., Park, H.: Weakly supervised nonnegative matrix factorization for user-driven clustering. *Data Mining and Knowledge Discovery* pp. 1–24 (2014)
6. Cochran, E., Lawrence, J., Kaiser, A., Fry, B., Chung, A., Christensen, C.: Comparison between low-cost and traditional MEMS accelerometers: a case study from the M7.1 Darfield, New Zealand, aftershock deployment. *Annals of Geophysics* 54(6) (2012)
7. D’Hondt, E., Stevens, M., Jacobs, A.: Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring. *Pervasive and Mobile Computing* 9(5), 681–694 (2013)
8. Ganti, R., Ye, F., Lei, H.: Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine* 49(11), 32–39 (November 2011)
9. Lee, B.T., Son, S.C., Kang, K.: A blind calibration scheme exploiting mutual calibration relationships for a dense mobile sensor network. *IEEE Sensors Journal* 14(5), 1518–1526 (May 2014)
10. Limem, A., Delmaire, G., Puigt, M., Roussel, G., Courcot, D.: Non-negative matrix factorization under equality constraints—a study of industrial source identification. *Applied Numerical Mathematics* 85, 1–15 (November 2014)
11. Lipor, J., Balzano, L.: Robust blind calibration via total least squares. In: Proc. of ICASSP. pp. 4244–4248 (May 2014)
12. Miluzzo, E., Lane, N., Campbell, A., Olfati-Saber, R.: CaliBree: A self-calibration system for mobile sensor networks. In: Proc. of DCOSS. LNCS, vol. 5067, pp. 314–331 (2008)
13. Plouvin, M., Limem, A., Puigt, M., Delmaire, G., Roussel, G., Courcot, D.: Enhanced NMF initialization using a physical model for pollution source apportionment. In: Proc. of ESANN. pp. 261–266 (2014)
14. Saukh, O., Hasenfratz, D., Thiele, L.: Reducing multi-hop calibration errors in large-scale mobile sensor networks. In: Proc. of IPSN (2015)
15. Saukh, O., Hasenfratz, D., Walser, C., Thiele, L.: On rendezvous in mobile sensing networks. In: Proc. of REALWSN. LNCS, vol. 281, pp. 29–42 (2014)
16. Schulke, C., Caltagirone, F., Krzakala, F., Zdeborova, L.: Blind calibration in compressed sensing using message passing algorithms. In: NIPS 26. pp. 566–574 (2013)
17. Sharp Corp.: GP2Y1010AU0F compact optical dust sensor (2006), datasheet
18. Wang, C., Ramanathan, P., Saluja, K.: Moments based blind calibration in mobile sensor networks. In: Proc. of ICC ’08. pp. 896–900 (May 2008)